WPS7747

Do Returns to Education Depend on How and Who You Ask?

Pieter Serneels
Kathleen Beegle
Andrew Dillon

WORLD BANK GROUP

Abstract

Returns to education remain an important parameter of interest in economic analysis. A large literature estimates returns to education in the labor market, often carefully addressing issues such as selection, into wage employment and in terms of completed schooling. There has been much less exploration of whether estimated returns are robust to survey design. Specifically, do returns to education differ depending on how information about wage work is collected? Using a survey experiment in Tanzania, this paper investigates whether survey methods matter for estimating mincerian returns to education. The results show that estimated returns vary by questionnaire design, but not by whether the information on employment and wages is self-reported or collected by a proxy respondent (another household member). The differences due to questionnaire type are substantial varying from 6 percentage points higher returns to education for the highest educated men, to 14 percentage points higher for the least educated women, after allowing for non-linearity and endogeneity in the estimation of these parameters. These differences are of similar magnitudes as the bias in OLS estimation, which receives considerable attention in the literature. The findings underline that survey design matters for the estimation of structural parameters, and that care is needed when comparing across contexts and over time, in particular when data is generated by different surveys.

Do returns to education depend on how and who you ask? [1] [2]

Pieter Serneels

University of East Anglia, IZA

Kathleen Beegle

The World Bank, IZA

Andrew Dillon

Michigan State University

Key words: returns to education, survey design, field experiment, development, Africa

JEL codes: J24, J31,C83

1. Introduction

Surveys remain a primary approach to empirical data collection in economics and the social sciences, yet variation exists in the design and protocols for implementing these surveys across study contexts. These discrepancies in survey design may induce different sources of non-random measurement error, but the potential biases are often left unquantified, which is particularly relevant when estimating structural parameters. This paper investigates whether survey methods matter for estimating returns to education, a common parameter in labor, development and public economics. Using a survey field experiment the study randomly allocates two common types of surveys, referred to as a long and a short questionnaire, and two common methods of response: response by the respondent him or herself, or by a proxy respondent. The results indicate that estimated returns depend on whether using a detailed or short questionnaire, but not on whether the respondent was interviewed herself or by proxy. The effect of questionnaire in our randomized control trial varies considerably by gender and level of education which underscores the potential heterogeneous effects of survey design bias across populations.

Empirical estimation of Mincerian returns to education using earnings functions (Mincer 1958) has become standard practice, and frequently involves comparison of returns over time, across countries and across sub-groups within countries (see for example Psacharopoulos and Patrinos (2004), Shultz (2004), or World Bank (2012), among others). This analysis often relies on data generated from different surveys. The comparisons commonly find differences that are in some cases large, and in other cases only a few percentage points apart. Much attention has been given to limitations of comparability due to differences in data sample coverage (selection issues) and method of analysis (non-linearity). A body of work also analyzes how to get structurally accurate ('true') estimates for returns to education, taking endogeneity into account.[3] These estimations provide important inputs to policy debates and decisions, especially in low and middle income countries where education attracts tremendous attention as a

---

[3] Alternative ways to estimate returns to education include instrumental variable techniques (Angrist and Kruger (1991), Card (2001), Cruz and Moreira (2005), among others), twin studies (Behrman et al. (1994, 1996), Miller et al. (1995), Bound and Solon (1999), among others) or randomized control trials. Our deliberate focus on cross section survey data arises from this being still the most common approach, especially in international comparisons, developing country settings, and education policy debates, as cross-country and time varying estimates remain costly to obtain.

contributor to economic growth and poverty reduction. However, little consideration has been given to possible discrepancies arising from differences in the survey method used.

A growing body of evidence indicates that survey methods matter for the labor statistics they generate. This is well illustrated for OECD countries, especially for the US (see Bound et al 2001), and recent evidence confirms this for developing countries (see Bardasi et al 2011).[4] Survey methods may impact estimates of structural parameters if measurement error varies with survey design in a systematic way. Non-random measurement error in a continuous left hand side variable would normally not bias OLS point estimates (although it may reduce precision).[5] However, it may lead to bias if the measurement error comes from respondents strategically trying to guess the true value of a variable on which they have imperfect information, and making a systematic error in this judgment. Response by proxy may result in this type of error as it is potentially prone to strategic guessing. A common example is that husbands may be systematically under-reporting earnings of their wives (or vice versa), for instance because they are not aware of all activities of the latter, and the earnings related to them.[6] Similarly, the wording and detail of questions may affect how respondents categorize themselves, and thus the labor statistics generated from the data. These differences, for instance in labor force participation and occupational distribution, may result in distinct subsamples on which estimations are carried out. This is especially relevant for returns to schooling, which are typically estimated for wage workers alone. Yet little is known how factors of survey design affect estimated returns to education.

To address this gap, this study implemented a field experiment that creates variation in two key dimensions of survey design: the level of detail of the questions - whether including or excluding employment screening questions - and the respondent selected: self or proxy – and investigates whether these differences in survey design yield distinct estimates of the returns to education using a common econometric approach and identification strategy.[7]

---

[4] While work on survey measurement error in high income countries often focuses on how survey replies deviate from true values using validation studies – comparing for instance in the case of wages between employee survey replies and employer pay records - this approach is typically less fruitful in low income countries where records are often missing, incomplete or less reliable. As a result, it is more fruitful to compare replies obtained from different survey methods. (See de Mel et al (2009) for a comparable approach to measure small business profit.)

[5] In contrast, in the case of discrete dependent variables, such as labor force participation, both classical measurement error and measurement error in the dependent variable biases the point estimates of the coefficients of right-hand-side variables (Hausman, Abrevaya, and Scott-Morton, 1998; Hausman, 2001.)

[6] This is especially relevant when husbands and wives have separate sources of income stemming from individual activities; which is often the case in both urban and rural developing country contexts.

[7] Note that the paper does not aspire to improving upon commonly used estimation strategies, but rather takes this as given, to focus on the impact of survey methods when using existing estimation methods.

The study results indicate that survey methods matter, in particular the use of short versus detailed questionnaire. While linear OLS estimates suggest that the short questionnaire generates significant differences for men but not for women, a more advanced analysis that allows for non-linear returns and accounts for endogeneity, finds significantly different results for the short questionnaire both in the case of men and women. The short questionnaire yields 6 percentage points higher returns to education for the highest (tertiary) educated men, and 14 percentage points higher returns for the lowest (primary) educated women. These discrepancies are of a similar or larger magnitude as commonly observed biases associated with simple OLS estimation which are the subject of a large literature (see for instance Card 1999); and therefore deserve attention. The divergences stem from differences in the categorization of subjects into wage work, caused by the absence of 'screening' questions in the short questionnaire. We observe no differences in estimated returns for proxy versus self-response. The results underline that care is needed when comparing estimates across surveys of different design. They also accentuate the importance of choosing the most appropriate design and safeguarding consistency in design across surveys, providing two simple but important guidelines for future empirical research.

The structure of the paper is as follows. The next section provides background and discusses relevant studies. Section 3 describes the experiment as well as estimation strategy. Section 4 provides a description of the data, while Section 5 presents the results. Section 6 concludes.

2. Background and literature

Mincer (1974) summarizes how returns to education can be estimated from a simple wage equation. Two methods are commonly used to achieve this, and in both cases the dependent variable is the log of wages. Estimation is typically carried out separately by gender due to differences in labor market opportunities for men and women. We focus on the approach that studies the effect of years of schooling (S) on the log of wages ($\ln W$), controlling for experience (E) and its squared term (E^2):

$$\ln W_i = \beta_0 + \beta_1 S_i + \beta_2 E_i + \beta_3 E_i^2 + \varepsilon_i \tag{1}$$

The coefficient of years of schooling reflects the average returns to education, representing the change in wages due to a change in years of schooling ($\beta_1 = \partial \ln w / \partial S$).[8] Psacharapolous and Patrinos (2004) provide an overview of returns to education estimates worldwide and over time using the Mincerian approach, and conclude that they are in the neighbourhood of 10%, with large variations around the mean. Their country specific estimates for men vary from zero returns for Italy [in 1989] to 28% for Jamaica [1989], and for women from negative returns of -0.8% for Suriname [1993] to 41% for Puerto Rico [1959].

Current and past work shows a general awareness that these comparisons suffer from a number of limitations. A first constraint is the difference in data sample coverage. While estimates for a country's average returns to education should be based on representative samples, this is not always the case. The estimation sample may contain only formal wage workers, excluding casual and informal wage workers. The data may be obtained from firm surveys, focusing on representativeness for a subset of firms rather than workers.[9] Other causes of concern relate to the method of analysis used. Coefficients obtained from the 'dummy variable' (second) approach need to be adapted before they can be compared to those obtained from the continuous approach presented above. There is also substantial variation in what is included as right hand side variables, as some models include occupational variables, leading to weaker returns, while others do not.[10] A key concern that has received much attention relates to the estimation method, as coefficients obtained from OLS are biased since they do not take endogeneity into account. IV or control function estimation addresses this concern provided they make use of valid instrumental variables (Blundell, Daerden and Sianesi 2005). Standard estimates also typically neglect heterogeneity in returns across different levels of education. Yet, investment in education may depend on whether returns are convex or concave, which is a subject of debate, including in Tanzania. While many of these concerns had long been neglected in analysis for developing countries, and especially for African countries (see Schultz 2004), they are now increasingly taken into account. [11]

---

[8] The alternative method replaces years of schooling by dummy variables for different levels of education, and obtains the returns to education by dividing the estimated coefficient by the corresponding years of schooling for each level of education relative to the level of education below.

[9] Firm survey often target a specific sector, for instance the manufacturing sector, or the private sector.

[10] Glewwe (1996) also suggests that substantial variability in school quality could bias returns to education and therefore should be included as right hand side variables. .

[11] Note that recent work indicates that the Mincerian model and its underlying assumptions do not necessarily hold in all contexts, even when addressing these concerns. Heckman, Lochner and Todd (2003) find empirical support for the model's predictions for the US for 1940s and 1950s, but not thereafter. Relaxing key assumptions seem to lead to different inference, and the authors conclude that Mincerian returns, in their context, do not necessarily provide good guidance to policy analysis. Caution is therefore needed, both in the interpretation of the results and

Surprisingly little attention has been paid to whether differences in survey methods may affect the estimated returns to education. This is especially relevant for low income countries, where analysis more often relies on data generated by different surveys. This study focuses on two dimensions of survey design (i) the specific questions used in the questionnaire, considering two common approaches - and (ii) whether the questions are answered by the respondent himself or herself, or by a another person.

The detail and wording of questions can have important effects on survey findings. In particular, labor questions have been shown to affect the labor statistics they generate, including those reflecting labor force participation and occupational distribution. This may be especially relevant in settings where there is substantial variation in the proportion of individuals who are employed in wage work versus household-owned enterprises or home production (who are not directly remunerated in the form of a salary or wage). Likewise, challenges may arise in settings where employment is highly seasonal or where an important proportion of workers are casual laborers. These differences in the categorization of subjects (in labor force participation and wage work) may also affect the estimated returns to education, as they result in a different subsample of wage workers for whom the returns are estimated.

Several studies have investigated different aspects of questionnaire design, including question style and wording (open vs. closed questions; positive vs. negative statements; etc.) or the specific place of questions within the survey questionnaire (see Kalton and Schuman (1982) for a review). While the general conclusion is that question-wording can have important effects, the direction of these effects is frequently unpredictable. Sustained research efforts to revise employment questions in the US provide interesting insights. Concerned that irregular, unpaid, and marginal activities may be underreported in the Current Population Survey (CPS), among others because people may not think of their activity as work, respondents were asked in a debriefing study to categorize different hypothetical situations as "work", "job", "business", and so on. While the majority were able to classify consistent with CPS definitions, large minorities gave incorrect answers for each vignette. Thirty eight percent of the respondents for instance

---

the advice to policy makers. Nevertheless this is still the dominant approach in the analysis of education and related policies, in particular in low income countries, and it is therefore useful to analyze their sensitivity to survey methods.

categorized non-work activities as "work" (Campanelli, Rothgeb, and Martin, 1989).[12] A 1991 experiment to evaluate the CPS questionnaire revision used direct screening questions and vignettes for unreported work and found that both the sequence of questions as well as their wording influenced respondent interpretation of work and affected the employment statistics generated (Martin and Polivka, 1995). Specifically, and of interest for our setting, the study found that using direct screening questions helped in detecting under-reporting of work related to household business or farm, as well as underreporting of teenage work. These concerns are highly relevant for developing countries, where much work activity is informal and risks to go undetected in a survey. This may be particularly true for female work, and several scholars have expressed concerns about the underreporting and undervaluing of women's work when using common survey methods to collect employment data (Anker, 1983; Dixon-Mueller and Anker, 1988; Charmes, 1998; Mata Greenwood, 2000).[13] In previous work, we illustrate that this is also relevant in the Tanzanian setting when obtaining labor statistics. Using a short questionnaire, without screening questions, seems to induce individuals to adopt a broad definition of employment, which tends to include domestic duties. But even after reclassifying these – to obtain the correct ILO classification, the short module results in lower female employment rates, higher working hours for both men and women who are working, and lower rates of wage work (Bardasi et al 2011).

Surveys with a labor content can be categorized into two groups: those making use of multiple detailed questions to find out whether the respondent is economically active and participating in the labor force during the reference period (typically the last seven days), and those using one question only to determine this. To illustrate the importance of this survey design feature in recently implemented surveys, Table 1 provides an overview of surveys with a labor content in sub-Sahara Africa for 2009-2012. Of the 21 surveys, 12 (57%) used a long and detailed questionnaire, while the remaining 9 (43%) used a short questionnaire. Our survey experiment implements short versus detailed questionnaires to reflect this variation.

Another dimension of survey design which may have implications on structural parameter estimation is the type of respondent. Different surveys adopt distinct approaches to choosing

---

[12] See Esposito et al. (1991) for methods to obtain diagnostic information in order to evaluate the effect of questionnaire revisions.

[13] Women often play a dominant role in household, domestic and agricultural activities, which may go unregistered as this is culturally often not considered as work (Mata Greenwood, 2000). The 1991 CPS study for the US also found gender dimensions of the survey effects, with the revised questionnaire trying to better capture unpaid work in a household business and farm activities, which increased the female employment rate.

who answers the questions. Household surveys in developing countries typically ask the head of the household, usually the male spouse, to answer employment questions about all household members.[14] Proxy respondents may not always provide accurate information, and this may bias the obtained statistics on employment and their distribution (Hussmanns, Mehran, and Verma, 1990). One alternative is to ask individual household members above a certain age directly; as do the Living Standards Measurement Study surveys (LSMS)[15] and Labor Force Surveys (LFS). However, requiring self–reporting from all individuals puts an extra logistical and financial burden on the fieldwork, and in practice survey managers often face a trade-off between information accuracy and the cost to obtain it.[16] An experimental study for the US investigated the potential bias of proxy versus self-response for health statistics, finding that randomly selected subjects reported fewer health events for themselves than for other household members (Mathiowetz and Groves 1985). In earlier analysis, we assess the implications of survey design for descriptive statistics and find important effects of proxy versus self-reporting on resulting labor statistics in Tanzania, observing that reporting by proxy leads to lower male employment rates, mostly due to underreporting of agricultural activities (Bardasi et al 2011). This bias is reduced when the proxy respondent are spouses or have some schooling. The consequences of these observed differences in categorization for the estimates of structural parameters, such as the returns to education, remain, however, unknown.

3. Experimental design and estimation strategy

Given the limited evidence of the effect of survey design choices on returns to education estimates, the survey experiment focused on the two key dimensions discussed above: (i) the level of detail of the questionnaire, more specifically the screening questions to establish employment status, and (ii) the type of respondent, namely self versus proxy response. Households were randomly selected and allocated to one of the four survey assignments based on these two dimensions. All household members aged 10 or above were eligible to respond to the individual-level module on employment, and we consider all labor force participants to estimate the returns to education.

---

[14] This is the approach followed by standard surveys like for instance Household Budget Surveys (HBS), Household Income/Consumption Expenditure Surveys (HICES) and Core Welfare Indicator Questionnaires (CWIQ) among others.

[15] See Grosh, and Glewwe (2000)

[16] Interviewing only self-respondents may require many re-visits, which can be costly. Response by proxy rather than individuals themselves reflects the common practice to interview an informed household member (often the household head or spouse), rather than each individual him or herself. In practice proxy respondents are often used when individuals are away from the household or otherwise unavailable in the time allotted in an enumeration area to conduct interviews.

To vary the detail of the questionnaire, we developed a short and a detailed labor module focusing on differences in screening questions used to determine economic activity and labor force participation. The detailed questionnaire reflects the approach that is generally considered to be best practice and is typically used in multipurpose household surveys, such as the Living Standard Measurement Surveys (LSMS). Increased demands from policy makers to evaluate changes over time, requires frequent data collection that is also simple to implement. Many countries therefore collect data on an annual basis using short questionnaires. The short module used in our survey experiment reflects the approach followed by more concise surveys used in many low-income countries, such as the Core Welfare Indicator Questionnaire (CWIQ) and the Welfare Monitoring Surveys (WMS), as well as other surveys listed in Table 1.

Specifically, the detailed module contains three questions at the start to determine employment status, namely, (i) whether the person has worked for someone outside the household (as an employee), (ii) whether s/he has worked on the household farm, and (iii) whether s/he has worked in a non-farm household enterprise. In each case the response is either yes or no.[17] In the short module there was only one question to determine employment status, namely whether s/he did any type of work, which also invited a response of yes or no. In both cases the questions were asked with respect to the last 7 days (the reference period for identifying those who are "employed" and the set of detailed questions on that employment) and, if not reported to work in the last 7 days, then asked for the last 12 months. Those identified as working in the last 7 days in either module were then asked identically the same questions to gather information on their occupation, sector, employer, hours, and wage payments in their main job. The short and detailed employment modules are reported in Annex. The analysis focuses on the response with respect to the last 7 days, which is generally considered the most reliable for labor market analysis.

To create variation in the second dimension, we randomly varied asking the questions directly to the respondent or a proxy respondent. The proxy respondent was randomly chosen among household members at least 15 years old to be able to compare across different proxy

---

[17] The short and detailed module in our experiment differed in two ways: in dropping the set of screening questions to determine employment status and in not asking about second and third jobs. Since we only consider labor outcomes in the first job, the analysis focusses on the effect of the additional screening questions.

respondent types.[18] The selected proxy then reported on up to two other randomly selected household members age 10 or older. Because in actual surveys, proxy respondents are not randomly chosen, but selected on the basis of availability, our experiment did not exactly mimic the actual conditions that result in proxy responses in household surveys.[19] However, it will provide estimates of proxy response bias[20] for a representative sample of potential proxies within the household.

The benchmark against which the commonly used approaches reflected by the short and proxy treatments are compared is the detailed self-report questionnaire, which is generally considered "best practice". Grosh and Glewwe (2000), providing detailed guidance to household surveys in developing countries recommend including screening questions. The use of multiple questions is also recommended by ILO as several categories of workers, including casual workers, unpaid family workers, apprentices, household members engaged in non-market production, and workers remunerated in-kind, have difficulties to correctly interpret a one off question about "any type of work" as referring to their situation (Hussmanns, Mehran, and Verma 1990). [21]

The aim of this paper is to investigate whether point estimates of returns to education vary depending on the survey designs, using existing methods of analysis - not necessarily to provide new or more accurate estimates of returns to education for Tanzania. The following equation provides a benchmark specification using OLS to estimate the returns to education:

$$\ln W_i = \beta_0 + \beta_1 S_i + \beta_2^j \left(S_i \times T_i^j \right) + \beta_3 T_i^j + \beta_4 X_i + \beta_5 D_i + \varepsilon_i \tag{2}$$

---

[18] The design of our survey was informed by assessing Tanzanian CWIQ 2006 data which indicated that the average Tanzanian household in this area has between two to three adults who could serve as a proxy with a minimum age of 15. Our sample households had 2.7 members 15 years and older on average.

[19] An alternative research design to assess the effect of proxy respondents would have been to interview two members of the household who report on their own labor activities and proxy report on the other. We did not implement such a design because it proved to be too difficult to ensure a proper implementation for a medium to large sample. After consultation with counterparts in Tanzania, we concluded that it would be difficult to assure that proxy and self responses would be independent and would remain unaffected by the knowledge that another household member reports on the same information, given the normally social nature of this setting. The specific concern was that the design (and open communication about this design within the village) would trigger either a coordinated response by household pairs and/or accommodation of response to other's expectations, which would introduce potentially much larger (unobserved) respondent biases.

[20] If desired we can also exploit the information we have about the relationship between the proxy respondent and the individual on whom the proxy informs to assess whether there are systematic response patterns that depend on the proxy chosen. Because of small household sizes, in actuality the spouse of the head of the household was in the vast majority (77%) of cases chosen, corresponding to the usual proxy respondent in surveys. Therefore, not surprisingly, our results are similar when limiting our sample for proxy respondents to spouses only (results not reported).

[21] All survey assignments received in addition to the labor module also five other modules: household roster, assets, dwelling characteristics, land, food consumption, and non-food expenditures. The questions followed the same sequence and the same phrasing and recall periods.

where the dependent variable is the log of daily wages, constructed as weekly earnings divided by days worked, T_i^j are the indicator variables for the respective treatments j= *short, proxy* (short versus detailed, and proxy versus self), and are included on their own, as well as interacted with years of schooling (S_i). X_i refers to age and its squared term, while D_i represents district indicator variables, which we include to increase precision[22]. We test whether the coefficient of the interaction term with each of the survey treatments j is significantly different from zero $\left(H_0 : \beta_2^j = 0\right)$; rejecting the null provides evidence for the effect of either questionnaire design or proxy reporting on the returns to school.

The literature emphasizes three sources of bias in OLS estimates of returns to education: non-linearity, endogeneity to do with completed schooling, and sample selection when focusing on wage workers only (or any subsample). We assess whether survey effects are still present when accounting for each of these. To address nonlinearity, we use a spline function allowing returns to vary across levels of education, and estimate:

$$lnW_i = \gamma_0 + g(S_i) + \left(g(S_i) \times T_i^j\right) + \gamma_4 T_i^j + \gamma_2 X_i + \gamma_3 D_i + \varepsilon_i'' \tag{3}$$

with $g(S_i) = \sum_{n=1}^{N} \gamma_n \cdot max\{S_i - \eta_{n-1}\}I_{in}$ and $I_{in} = 1[\eta_{n-1} \leq S_i < \eta_n]$, with η_n the place of the n-th node for $n=1,2,...,N$. We consider two nodes, which we fix at 8 and 12 years of education respectively; this reflects the Tanzanian education system, where primary school is seven years, secondary school lasts four years, after which one can move to higher education. [23] γ_n then reflects the returns to schooling for the n-th interval. Returns are linear if $\gamma_1 = \gamma_2 =, ..., = \gamma_n$. Like before we then test whether the interaction of the returns to education and the survey treatments are significant.[24]

The concern of endogeneity of schooling is often addressed using instrumental variable estimation, while the control function approach is applied as an alternative, especially in the case of non-linear returns. Using a control function approach we estimate:

$$lnW_i = \delta_0 + g(S_i) + \left(g(S_i) \times T_i^j\right) + \delta_1 T_i^j + \delta_2 X_i + \delta_3 D_i + \delta_4 \hat{c}_i + v_i \tag{4}$$

and test whether the interaction terms are significant. The control function term $\hat{c}_i$ is obtained from the first stage estimation:

$$S_i = \pi_0 + \pi_1 Z_i + \pi_2 X_{2i} + \pi_3 D_i + c_i \tag{5}$$

---

[22] The results in the subsequent sections are robust to the in- or exclusion of these district indicator variables.
[23] Using different nodes leads to similar results.
[24] For other work using a related approaches see Schady (2003) and Soderbom et al (2010).

with education supply characteristics of the local community as identifying instruments Z_i, correlated with schooling but not with the unexplained variation in wages ε_i''.[25] To isolate the community specific effects, we also include another community characteristic, namely distance to an all-weather road.

Because the key difference between the long and short questionnaire is that the first includes employment screening questions, which may lead to a different sample of wage workers on which the estimates are carried out, an issue of special interest is whether controlling for a treatment specific selection correction term would further improve results. Following the classic Heckman approach we include the Inverse Mills ratio obtained from a first stage equation that models selection into wage work, and estimate:

$$lnW_i = \delta_0' + g(S_i) + \left(g(S_i) \times T_i^j\right) + \delta_1'T_i^j + \delta_2'X_i + \delta_3'D_i + \delta_4'\hat{c}_i + \delta_5'\lambda_i + v_i' \qquad (6)$$

Where the selection term (λ_i) is obtained in the usual way from:

$$P_i = \theta_0 + g(S_i) + \theta_1 T_i^j + \theta_2 X_{2i} + \theta_3 Z2_i + \theta_4 D_i + v_{2i} \qquad (7)$$

with P_i reflecting whether the individual is a wage worker or not, and variables contained in the selection equation but not in the wage equation include marital status and the number of dependents in the household.[26]

While in theory the model is fully identified when using the same variables in the first and second stage, identification then relies entirely on functional form (i.e. the non-linearity of the selection equation) and may be fragile. We follow common practice to include at least one additional identifying variable in the selection equation. While good instruments may be difficult to find, there are some valuable candidates, and we follow existing practice.[27] The early literature on labor supply includes family formation variables like marital status and number of children in the selection (but not in the wage) equation. Later work for the US and few other high income countries shows that marital status can affect male and in some cases female earnings. Such evidence is, however, absent for developing countries, which offer a very different

---

[25] Card (2001) dsicusses other work using variation in supply of education to identify causal effects.

[26] Addressing both the concerns of endogeneity related to attained schooling and that of sample selection due to estimation on a subsample (of wage workers) is receiving increased attention, see for instance Kuepie et al 2010. Studies that exploit exogenous variation in the supply of education, have also drawn renewed attention to the importance of the latter – see for instance Duflo (2001).

[27] We investigated the best candidate instruments in this context, including education reforms. One such reform was implemented in the late 1960s and changed the structure of the education system (Kerr 2011), another more recent policy change relates to the introduction of Univeral Primary Education (UPE) (Hoogeveen and Rossi, 2011). Both of these changes disqualify from being good instruments for our purpose, as the former is too old while the latter is too recent. No other changes in the organization of education have been identified for Tanzania.

context.   Indeed marital status is said to affect earnings through two specific channels: specialization and selection (Korenman and Neumark 1998). Regarding the former the argument is that marriage can increase (especially the husband's) specialization, leading to higher productivity and earnings. The selection argument states that more productive workers – who thus also have higher earnings – are more likely to find a partner and get married in the first place. Both of these seem to be largely absent in rural and provincial developing settings, such as the one in our sample, where gender roles are strong and marriage is often decided during teenage years.[28] Although we cannot entirely exclude that unobserved characteristics set at young age drive both marital status and productivity. Evidence for an effect of fertility on male and female earnings is also scarce for developing countries. Piras and Ripani,(2005) in a comparison of four countries in Latin America find little evidence that mothers earn lower wages than women with no children. McCabe and Rosenzweig (1976) consider the possible effects of fertility on wages, and argue explicitly that this depends on the compatibility of the specific occupation with child rearing. Our focus is on wage work, which is incompatible with child rearing, and the number of children is expected to affect the selection into wage work, but not on the job productivity.

In contrast, there is strong evidence that family formation - including marriage and fertility – has important effects on time use where markets for household chores and child care are missing. Being married increases the time devoted to housework, in particular for women, while the presence of children, especially small children, increases time devoted to care for both men and women (see World Bank 2012).   In this context, family formation affects primarily the probability of being in wage work, i.e. outside the informal sector or home work, and this is confirmed by the data, as discussed in Section 5.[29]

---

[28] The specialisation argument is that married men have more time to specialise in professional activities, but it is unclear whether this is the case in rural societies, where unmarried men (and women) often stay with their parents and hence do not have to carry out the household chores associated with an independent household. Similarly, marriage is often decided before labor market performance has been revealed, breaking the reverse causality that is a major concern in high income countries.

[29] Note that in the context where the survey experiment was implemented wage work mostly reflects formal sector work, while informal sector work is typically self-employed. Occupations like domestic worker – which are scarce compared to other settings such as for instance India – are self-employed rather than wage work. These variables are also commonly included in the early literature using selection equations, including for women (See for instance Mroz (1987)). For reasons of consistency we include the same variables in the selection equations for men and women

We include as identifying variables both marital status and the number of children.[30] A placebo test that includes these family formation variables in the second stage equation confirms that they have no effect on earnings in our setting.

A challenge with the above approach arises when the education variables determine occupational sorting, indicated by their statistical significance in the selection equation. This may lead to biased estimation results as the selection correction term is now correlated with the error term in the second stage. To address this we follow a procedure similar to the one applied by Duflo (2001) and originally suggested by Heckman and Hotz (1989).[31] The procedure exists of including the instrumental variables that are used to address the endogeneity of education (Z), i.e. the community distance variables used in the control function, in the selection equation, and include polynomials of the predicted probabilities of becoming a wage work ($\hat{p}$) in the main equation.

$$lnW_i = \delta_0'' + g(S_i) + \left(g(S_i) \times T_i^j\right) + \delta_1'' T_i^j + \delta_2'' X_i + \delta_3'' D_i + \delta_4'' \hat{c}_i + \delta_5' h(\hat{p}_i) + v_i'' \qquad (8)$$

Where the selection correction term $(\hat{p}_i)$ is obtained from:

$$P_i = \theta_0' + g(S_i) + \theta_1' T_i^j + \theta_2' X_{2i} + \theta_3' Z_{2i} + \theta_4' Z_i + \theta_5' D_i + v_{2i}' \qquad (9)$$

with P_i reflecting whether the individual is a wage worker or not, and Z2 variables contained in the selection equation but not in the wage equation include marital status and the number of dependents in the household, while Z reflect the community distance variables. As before, to isolate the community effects, we also include mean distance to an all-weather road.

4. Data and Context

The survey experiment was implemented in Tanzania, which has different types of labor market surveys, including CWIQs, LFSs and multipurpose household surveys, like the Household Budget Survey (HBS). These different data sources have been variably used to estimate returns to education. The experiment we conducted was the Survey of Household Welfare and Labour in Tanzania (SHWALITA). The field work was conducted from September 2007 to August 2008 in villages and urban areas from 7 districts across Tanzania: one district in the regions of Dodoma, Pwani, Dar es Salaam, Manyara, and Shinyanga region and two districts in the Kagera region. Households were randomly drawn from the listing of villages (urban clusters) and randomly

---

[30] Results are similar results if we separate out young (minus 6) and older children (6-15) – results not reported
[31] Existing work rarely addresses this, although this is receiving increased attention. On-going work investigates whether this is an important source of bias, see for instance Schwiebert (2015), revisiting Mulligan and Rubinstein (2008)

allocated to one of the four survey assignments. The total sample is 1,344 households, with 336 households assigned to each of the four survey assignments. Although the sample of 1,344 is not designed to be nationally representative of Tanzania, the districts were selected to capture variations between urban and rural areas and along other socio-economic dimensions.

The basic characteristics of the sampled households generally match the nationally representative data from the Household Budget Survey (2006/07) (results not presented here). Household interviews were conducted over a 12-month period, but because of small samples, we do not explore the survey assignment effects across seasons (such as harvest time with peak labor demand and dry seasons with low demand). The random assignment of households is validated when examining different household characteristics, as reported in Table 2 Panel A.

Individuals are classified on the basis of the survey assignment they actually received, which is the result of the initial assignment of their household (to one of the four survey assignments), whether the individual is selected to be a proxy respondent or a self-report, and whether the proxy/self-report assignment is realized. In the case of self-report modules, up to two persons over age 10 are randomly selected to self-report. If persons randomly selected to self-report are unavailable, an alternative person is selected at random. In the case of proxy assignment, one person in the household over the age of 15 is selected to self-report and to proxy report on up to two random household members. Thus, in the proxy assignment, one household member actually self-reports in addition to reporting on other household members. Therefore, the number of self-reports should be about half the number of proxy reports for households in the proxy assignment. In total, by design, there are more self-reports than proxy reports. Because the survey experiment highly emphasized the importance of avoiding proxies when not assigned to this treatment, the project was successful at completing self-reports when assigned. In about five percent of cases, the team was unable to interview a person selected for self-report. While there were small deviations from the original design during the implementation, the overall realised design remained very close to the planned design, as shown in Table A.2. in annex. The results presented in this paper are unchanged if we exclude the observations where we had to deviate slightly from the planned design, or when we restrict ourselves to spouse instead of random proxies to reflect the more common approach in household surveys. Panel B of Table 2 reports balance tests at the individual level, and shows that allocation across treatments is generally well balanced. There appear to be some imbalances across individuals related to age,

marital status and number of children, but this applies to the proxy assignment, and not to the short – detailed assignment, where characteristics are balanced.

The identifying variables for the control function estimation are obtained from CWIQ data collected in the same communities during the same year, but covering different households. We use the community mean distance to primary and secondary school to proxy the local supply of education at the time of the individual's schooling.

Before discussing the results, we consider the differences in wages across treatments. The plots in Figure 1 present the kernel density for the log of daily earnings for the different subsamples. Figure 1A(i) and 1B(i) draw the earnings obtained from the detailed and short modules for men and women respectively, with the distribution generated by the latter mostly to the left of that generated by the former. Figure 1A(ii) and 1B(ii) suggests that the difference is less outspoken for the wages generated by proxy and self treatments.

The descriptive statistics in Table 3 reveal further differences. While the detailed module yields a lower proportion of labor force participants compared to the short module for both men and women, it generates a higher relative proportion of wage workers for both sexes. The detailed module also produces lower mean daily earnings than the short module. Proxy response, on the other hand, leads to relative lower labor force participation and a lower proportion of wage workers, but produces higher mean wages compared to self response, again for both sexes. Although the differences in wages between treatments may seem small, they are actually substantial, as is perhaps more clear when expressed as monthly rather than daily wages. The average daily wage reported in Tanzanian Shilling (TSH) in Table 1 for men (3987, 4781, 3634, 6255) correspond to monthly income expressed in USD of respectively 75USD, 89USD, 68USD and 117USD, while the respective daily wages in TSH for women (3912, 4388, 3552, 5367) correspond to the monthly wages of 73USD, 82USD, 66USD, 100USD.

5. Analysis and results

As a benchmark, we estimate equation (2) using OLS, which assumes linearity and does not account for endogeneity.[32] The results are reported in Table 4 and suggest that average returns to education in our sample are 8% for men and 10% for women (See Columns 1 and 4). Standard errors are clustered at the household level. These results remain when including survey treatment variables (Columns 2 and 5). Including interaction effects, the results in Column 3 indicate that for men the short module yields substantially and significantly higher returns to education, but proxy response does not affect returns to education. For women interaction effects with either the short or the proxy treatment are insignificant, as shown in Column 6.

We proceed by estimating the models that allow for non-linearity and endogeneity. Existing evidence indicates that returns to education are often non-linear, in particular for Tanzania (see Soderbom et al (2010); Kerr (2011)). This is confirmed for our data as shown by the plot of a non-parametric kernel regression in Figure 2, which suggests an *S*-shaped pattern for men and convexity for women.

We estimate equations (3) and (4) presented in Section 3 to obtain nonlinear returns while allowing for endogeneity using a control function approach. The results reported in Columns 1-3 of Table 5 for men and Columns 4-6 for women show strong nonlinearities for both men and women. Returns tend to increase with the level of education, but at a decreasing rate. This is consistent with other results for Tanzania (see Kerr (2011), Soderbom, et al (2006)), but is less outspoken for women once we include interaction terms between treatment and years of education.

Column 3 and 6 demonstrate that the interaction effect between the short survey treatment and level of education occurs at tertiary education for men and both primary and tertiary education for women.[33] The (gender specific) control function term is substantial in size and its inclusion affects the estimation results. As expected, control function estimates of the returns to education

---

[32] An alternative approach would be to carry out separate estimation per treatment group and this yields the same results. Because of the small sample size we focus on the pooled results.

[33] Additional testing shows that the interaction effects with the short treatment are jointly significant for women but not for men, while interaction effects with proxy are not jointly significant.

are larger than OLS estimates indicating confirming that the latter are biased due to unobserved characteristics.

Table 6 reports the first stage estimates for the control function and shows the importance of the identifying variables. As expected the community mean distance of secondary school is especially important for men, while both the community mean distance to primary and secondary school are of particular relevance for women. To avoid that these variables proxy for community fixed effects we also include another community characteristic reflecting general isolation, namely distance to the nearest all-weather road. This variable is then also included in the second stage regression. An F-test for joint significance of the instruments yields p-values at the 0.06 for men and below 0.01 for women.   In a placebo test where we add the two identifying variables to the second stage regression (but without the control function term), the parameter estimates are not significant.

The next step is to estimate equations (6)-(9) to correct for selection into wage work. Table 8 presents the results for the selection equation. Two issues stand out. First, education has a significant effect on the probability of being a wage worker for both men and women.[34] As discussed in Section 3, this raises an additional challenge, and leads us to compare with an alternative approach. Second, the coefficients of the treatments indicate that using a detailed or short questionnaire has strong effects on who is categorised as wage worker, both for men and women, while the effects are less strong and less robust for proxy versus self-response. The effect of the short questionnaire reflects its key difference from the detailed module, which includes additional screening questions at the beginning of the questionnaire. Omitting these questions seems to lead to a different categorization of respondents into wage work, and as a result the wage equation is estimated on a different sample. The descriptive statistics in Table 3 already indicated that both labor force participation and especially the proportions of wage workers differ substantially between the detailed and short questionnaires.

Column 1 and 3 in Table 8 present the estimates of equation 7 for men and women respectively, reflecting the traditional Heckman first stage that includes Z2. The number of children is significant with a p-value of 0.02 for men, while being married has a p-value of 0.11 for women. The instruments are jointly significant at 0.2 for men and 0.06 for women. A placebo test, where

---

[34] We allow for non-linearity to maintain consistency between first and second stage equations. Descriptive statistics as well as kernel density regression also suggest non-linear education effects for the selection into wage work (not reported).

the family formation variables are included in the wage equation confirm that they have no significant effect on wages. Columns 2 and 4 in Table 8 present the estimates of equation (9), which includes both sets of instruments, namely Z1, the family formation variables, and Z, the community school variables used in the earlier part of the analysis.

Table 7 presents the second stage estimation results, with Columns 1 and 3 reporting the estimates from the classic Heckman approach, relying on first stage presented in Table 8 Columns 1 and 3, while Columns 2 and 4 report the estimates from the alternative Heckman-Hotz approach that relies on the first stage presented in Table 8 Columns 2 and 4. Both sets of results confirm that male returns to education are higher among tertiary educated when using the short questionnaire, while returns for women are higher for both the primary and tertiary level when using the short module.

Table 9 summarizes our estimates for the returns to education across estimation methods, focusing on the difference in between detailed and short questionnaire, as the difference with proxy treatment is not significant. While it is difficult to make bold claims, given the sample sizes, the message is consistent across estimation methods despite the small sample size. Using the detailed questionnaire with self-response as the best practice reference[35], the results indicate that the short questionnaire systematically overestimates the returns for men, especially for tertiary educated men and women, and primary educated women. The preferred models, which also account for selection, confirm this bias, with estimated returns for tertiary educated men and women 6 and 8 percentage points higher respectively when using the short module, and 14 percentage points higher for primary educated women. In light of the existing debate concerning the convexity of returns to education in developing countries, and especially in Tanzania, these results suggest that this convexity may have been overestimated when studies use data obtained from a short questionnaire.

These results make the general point that questionnaire design can have both significant and substantial effects on the estimation of structural parameters. That we find these effects – despite the small sample size of our data, provides strong evidence that survey design matters particularly in estimating male and female returns to education across different levels of schooling. The findings highlight the need for caution when comparing structural estimates

---

[35] As described above, following existing work we consider the detailed questionnaire as best practice .

obtained from data generated by different survey methods. At a practical level, they also underline the importance of consistency and best practice in survey design.

6. Conclusion

This paper investigates whether survey design matters for estimating returns to education using a field experiment, and finds that it does. Using a randomized intervention that implemented two variations of survey design, namely use of a commonly used short versus a detailed labor questionnaire, and self-response compared to response by proxy, we find that estimated returns to education differ dependent on the survey instrument, but not on the type of respondent, for both men and women. The short questionnaire leads to biased estimates of returns to education relative to the detailed questionnaire. The biases are substantial and significant, resulting in higher estimates in the short questionnaire, ranging from 6 and 8 percentage points higher for tertiary educated men and women respectively, to 14 percentage points higher for primary educated women. These results are robust when accounting for non-linearity of education and taking both endogeneity of education and selection into wage work into account, making use of commonly applied estimation and identification methods. The results are consistent with suggestive evidence from qualitative research, including respondent debriefing studies in the US which show that screening questions can have important effects on the labor statistics they generate.

These observed differences are of a similar magnitude as the estimation bias related to endogeneity which is the subject of considerable attention in the literature; they are also of a similar magnitude as the differences in estimated returns between gender, levels of education, across sectors (public, formal private and informal private sectors) observed in the region, and therefore deserve attention.[36]

While this paper does not aspire to obtaining more accurate estimates of returns to education for Tanzania - the data was not collected with this aim in mind and using standard questionnaires, and the sample is also small and not nationally representative - it is useful to consider the existing results for Tanzania, in particular because they rely on data from different surveys. Nerman and Owens (2010), using the 2001 and 2007 waves for the nationally representative

---

[36] A review of key overview papers for the region indicates differences between 2 and 18 percentage points across these dimensions (see Teal and Baptist 2014; Schultz 2003; Kuepie et al 2009).

Household Budget Surveys estimate returns to education in order to explain demand for education and report OLS estimates between 0.3% to 16.3%, depending on the sub-group, and not controlling for endogeneity. Using nationally representative cross-section data from of the Integrated Labour Force Surveys (IFLS) for 2001 and 2006, Kerr (2011) estimate returns between 8% and 13% when using OLS. When allowing for nonlinearities the results suggest that returns are strongly convex, but when also addressing endogeneity exploiting a change in the education system in the mid 1960s, returns are concave and higher at the lower levels of education, which the authors argue reflects an ability bias.[37] These results also shed light on earlier findings by Söderbom, Teal, Wambugu, Kahyarara (2006), who, using data for employees in the manufacturing sector in Tanzania for 1993, 1994, 1999 and 2001, also find a convex earnings function when taking endogeneity into account. Their estimates exceed the OLS estimates, which may be a consequence of self-selection on ability into the manufacturing sector. Our estimates also aligh well with those from other African countries. Schultz 2004 reports wage gains of 5-20% for each year of schooling for five African countries.[38]

Our results underline that survey methods matter for the estimation of structural parameters, such as Mincerian returns to education, and indicate that care is needed in the comparison of these returns across data sources, as is typically the case with worldwide comparisons, as well as comparisons over time

7. References

Angrist, J. and A. Krueger. 1991. "Does Compulsory School Attendance Affect Schooling and Earnings?" *The Quarterly Journal of Economics* 106(4): 979-1014.

Anker, R. 1983. "Female Labour Force Participation in Developing Countries: A Critique of Current Definitions and Data Collection Methods." *International Labour Review* 122(6):709-724.

Bardasi, Elena, Kathleen Beegle, Andrew Dillon, and Pieter Serneels. 2011. "Do Labor Statistics Depend on How and to Whom the Questions Are Asked? Results from a Survey Experiment in Tanzania." *World Bank Economic Review* 25(3): 418-447.

---

[37] While the instrument reflects an exogenous policy change that extended primary school from 4 to 7 years in the mid 1960s, it is unclear to what extent it is applicable to the entire sample of analysis. As the change in policy took place 40 years ago, it had an immediate impact on those who were 10-13 years old in the mid 60s, or 50-53 in the mid 2000s. With average age of 38, this is likely to be a minority in the IFLS sample. Because the change in policy took place so long ago, we do not consider this as identifying variable for our control function estimates.
[38] Kenya, Nigeria, Ghana, Burkina Faso, and Cote d'Ivoire

Battistin E. and B. Sianesi. 2011. "Misclassified treatment status and treatment effects: an application to returns to education in the United Kingdom." *Review of Economics and Statistics* 93(2): 495–509

Behrman, Jere R., Mark Rosenzweig, and Paul Taubman. 1994. "Endowments and the allocation of schooling in the family and in the marriage market: the twins experiment." *Journal of Political Economy* 102 (6): 1134–1174.

Behrman, Jere R., Mark Rosenzweig, and Paul Taubman. 1996. "College choice and wages: estimates using data on female twins." *Review of Economics and Statistics* 73 (4): 672–685.

Biemer, P., R. Groves, L. Lyberg, N. Mathiowetz, and S. Sudman. 1991. *Measurement Error in Surveys.* New York: John Wiley & Sons.

Bingley P., K. Christensen, and I. Walker. 2005. "Twin-based Estimates of the Returns to Education: Evidence from the Population of Danish Twins." mimeo.

Blundell, R., L. Dearden, and B. Sianesi. 2005. "Evaluating the Effect of Education on Earnings: Models, Methods and Results from the National Child Development Survey," (with L. Dearden and B. Sianesi) *Journal of Royal Statistical Society Series A,*

Blundell, R., M. Costa-Dias. 2009 "Alternative Approaches to Evaluation in Empirical Microeconomics," , *Journal of Human Resources, 44(3), 565-640*

Bommier, A. and S. Lamber. 2000. "Education Demand and Age at School Enrollment in Tanzania." *The Journal of Human Resources* 35(1).

Bound, J., C. Brown, and N. Mathiowetz. 2001. "Measurement Error in Survey Data." In *Handbook of Econometrics* Vol. 5. ed. J. Heckman and E. Leamer. Amsterdam: North-Holland, Elsevier Science.

Bound, John, and Gary Solon 1999. "Double trouble: on the value of twins-based estimation of the return to schooling." *Economics of Education Review* 18 (2): 169–182.

Burke, K., and K. Beegle. 2004. "Why Children Aren't Attending School: The Case of Northwestern Tanzania." *Journal of African Economies* 13(2): 333-355.

Campanelli, P., J.M. Rothgeb, and E.A. Martin. 1989. *The Role of Respondent Comprehension and Interviewer Knowledge in CPS Labor Force Classification.* American Statistical Association Proceedings (Survey Research Methods Section).

Card, D. 1999. "The Causal Effect of Education on Earnings." In *Handbook of Labor Economics* Vol. 3A. Ed. O. Ashenfelter and D. Card, Amsterdam: North Holland, Elsevier Science

Card D. 2001. "Estimating the Return to Schooling: Porgress on Some Persistent Econometric Problems." *Econometrica* 69(5): 1127-1160.

Carneiro, P., J. J. Heckman, and E. Vytlacil. 2010. "Estimating marginal returns to education." Working paper CWP29/10, Institute for Fiscal Studies, Department of Economics, University College London

Charmes, J. 1998. *Women Working in the Informal Sector in Africa: New Methods and New Data.* Paris: Scientific Research Institute for Development and Co-operation.

Cruz, L and M. Moreira. 2005. "On the Validity of Econometric Techniques with Weak Instruments: Inference on Returns to Education Using Compulsory School Attendance Laws." *Journal of Human Resources* 40 (2): 393-410.

de Mel, S., D. McKenzie, and C. Woodruff. 2009. "Measuring Microenterprise Profits: Must we ask how the sausage is made?" *Journal of Development Economics* 88(1): 19-31.

Dickson, M, and S. Smith. 2011. "What Determines the return to education: an extra year or hurdle cleared?" Centre for Market and Public Organisation, Working Paper No. 11/256, University of Bristol

Dixon-Mueller, R., and R. Anker. 1988. *Assessing Women's Economic Contributions to Development.* Training in Population, Human Resources and Development Planning Paper number 6, Geneva: International Labour Office.

Dohmen T., D. Jaeger, A. Falk, D. Huffman, U. Sunde, H. Bonin, (2010). Direct Evidence on Risk Attitudes and Migration, *Review of Economics and Statistics*, 92(3) 684–689.

Duflo E., 2001, Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment. American Economic Review Vol 91, No 4 (Sep 2001)

Esposito, J.L., P.C. Campanelli, J. Rothgeb, and A.E. Polivka. 1991. "Determining which Questions Are Best: Methodologies for Evaluating Survey Questions." *Proceedings of the American Statistical Association* (Survey Research Methods Section).

Grosh M. and P. Glewwe(eds). 2000. *Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 Years of the Living Standards Development Study*. Oxford University Press (for the World Bank).

Hausman, J.A. 2001. "Mismeasured Variables in Econometric Analysis: Problems from the Right and Problems from the Left." *Journal of Economic Perspectives* 15 (4): 57-67.

Hausman, J.A., J. Abrevaya, and F.M. Scott-Morton. 1998. "Misclassification of the Dependent Variable in a Discrete Response Setting." *Journal of Econometrics* 87: 239-269.

Heckman J.J., L. J. Lochner, and P. E. Todd. 2003. "Fifty years of Mincer earnings regressions." IZA Discussion Paper 775

Heckman J.J., V.J. Hotz,, 1989, Chooisng Among Alternative Nonexperimental Methods for Estimating the Impact of Social Programs: The Case of Manpower Training, Journal of the American Statistical Association, Vol 84, Issue 408 (Dec 1989)

Hill, D.H. 1987. "Response Errors in Labor Surveys: Comparisons of Self and Proxy Reports in the Survey of Income and Program Participation (SIPP)." In *Proceedings of the Bureau of Census, Third Annual Research Conference*.

Hoogeveen, J. and M. Rossi. 2011. "Saving Goat and Cabbages? Enrollment and grade achievement after the introduction of free primary education in Tanzania." mimeo

Hussmanns, R., F. Mehran, and V. Verma. 1990. *Surveys of Economically Active Population, Employment, Unemployment and Underemployment: An ILO Manual on Concepts and Methods*. ILO: Geneva.

Hyslop, D.R., and G.W. Imbens. 2001. "Bias from Classical and Other Forms of Measurement Error." *Journal of Business & Economic Statistics* 19(4): 475-481.

Jensen, R. 2010. "The (Perceived) Returns to Education and the Demand for Schooling." *Quarterly Journal of Economics*, May 2010.

Judge, G., and L. Schechter. 2009. "Detecting Problems in Survey Data Using Benford's Law." *Journal of Human Resources* 44(1): 1-24.

Kalton, G., and H. Schuman. 1982. "The Effect of the Question on Survey Responses: A Review." *Journal of the Royal Statistical Society* 145(1): 42-57.

Kasprzyk, D. 2005. "Measurement error in household surveys: sources and measurement." In *Household Sample Surveys in Developing and Transition Countries*, New York : United Nations, pp. 171-198.

Kerr, A. 2011. Estimating the Returns to Education in Tanzania, Chapter 2, PhD Dissertation, Oxford University, Economics Department

Koop, G., and J.L. Tobias. 2004. "Learning about heterogeneity in returns to schooling." *Journal of Applied Econometrics* 19: 827-849

Kuepie M., C.J. Nordman, F. Roubaud, 2009, Education and earnings in urban West Africa, Journal of Comparative Economics 37, 491-515

Martin, E., and A.E. Polivka. 1995. "Diagnostics for Redesigning Survey Questionnaires: Measuring Work in the Current Population Survey." *Public Opinion Quarterly* 59: 547-567.

Mata Greenwood, A. 2000. *Incorporating Gender Issues in Labour Statistics*. Geneva: International Labour Office, Bureau of Statistics.

Mathiowetz, N.A., and R.M. Groves. 1985. "The Effects of Respondent Rules on Health Survey Reports." *American Journal of Public Health* 75(6): 639-644.

McCabe, J. L. And M.R. Rosenzweig, 1976, Female labor-force participation, occupational choice, and fertility in developing countries, Journal of Development Economics, Elsevier, vol. 3(2), pages 141-160

Meghir, C. and S. Rivkin. 2010. "Econometric Methods for Research in Education." Institute for Fiscal Studies Working Paper W10/10

Mincer, Jacob. 1958. "Investment in Human Capital and Personal Income Distribution." *Journal of Political Economy* 66(4): 281-302
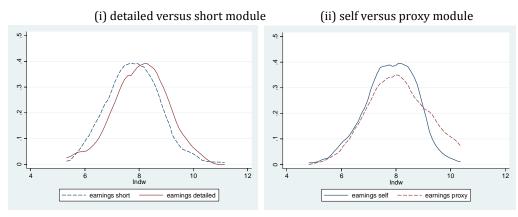
Mincer, Jacob. 1974. *Schooling, Experience and Earnings*. New York: Columbia University Press.

Miller, Paul W., Charles Mulvey, and Nick Martin. 1995. "What do twins studies reveal about the economic returns to education? A comparison of Australian and U.S. findings." *American Economic Review* 85 (3): 586–599.

Moore, J. 1988. "Self/Proxy Response Status and Survey Response Quality: A Review of the Literature." *Journal of Official Statistics* 4: 155-172.

Mroz, Thomas A. 1987. "The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions." *Econometrica* 55(4): 765-99.

Mulligan C.B., Rubinstein Y., 2008, Selection, investment and women's relative wages over time. The Quarterly Journal of Economics, 123(3):1061-1110

Nerman, M. and T. Owens. 2010. "Determinants of Demand for Education in Tanzania: Costs, Returns and Preferences." University of Gothenburg, Working Papers in Economics, No 472.

Piras C. and Ripani L., 2005, The Effects of Motherhood on Wages and Labor Force Participation: Evidence from Bolivia, Brazil, Ecuador and Peru. Sustainable Development Department Technical Papers Series

Psacharopoulos, George, and Harry Patrinos. 2004. "Returns to investment in education: a further update." *Education Economics* 12(2): 111-134.

Rankin, N., J. Sandefur, and F. Teal. 2010. "Learning and Earning in Africa: Where are the Returns to Education High?" Centre for the Study of African Economies Working Paper, 2010-02

Rosenzweig, M. R. 2010. "Microeconomic Approaches to Development: Schooling,Learning, and Growth." Yale Economic Growth Center Discussion Paper No. 985, 2010.

Schultz, P.T. 2004. "Evidence of Returns to Schooling in Africa from Household Surveys: Monitoring and Restructuring the Market for Education." *Journal of African Economies* 13 Suppl.2: ii95-ii148.

Schady, N. R. 2003. "Convexity and Sheepskin Effects in the Human Capital Earnings Function: Recent Evidence for Filipino Men." *Oxford Bulletin of Economics and Statistics* 65(2): 171-196.

Schwiebert J, 2015, Revisiting the Composition of the Female Workforce - A Heckman Selection Model with Endogeneity, mimeo

Soderbom, M., F. Teal, A. Wambugu, and G. Kahyarara. 2006. "The Dynamics of Returns to Education in Kenyan and Tanzanian Manufacturing." *Oxford Bulletin of Economic and Statistics* 68(3): 261-288.

Teal J.F., S. Baptist, 2014, Technology and Productivity in African Manufacturing Firms, World Development Vol. 64, pp. 713–725

World Bank, 2012, World Development Report 'Gender Equality and Development'.

8. Figures and Tables

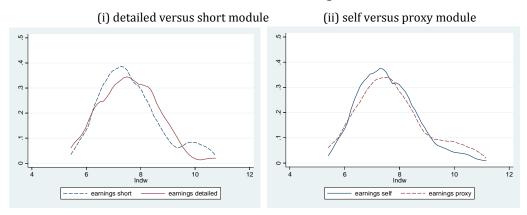### Figure 1: Kernel density of earnings by treatment

A. Male earnings

(i) detailed versus short module   (ii) self versus proxy module

B. Female earnings

(i) detailed versus short module   (ii) self versus proxy module

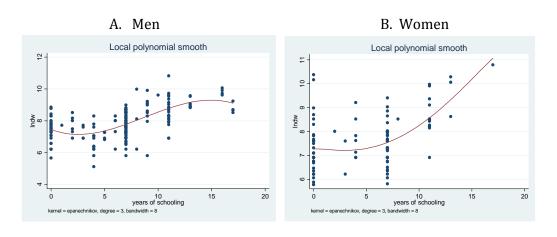### Figure 2: Kernel regression plot for wage equation

A. Men   B. Women

Table 1: Overview of types of recent surveys for the sub-Sahara Africa Region

Country	survey name	date	type of survey
Botswana	BCWIS	2009	detailed
Cameroon	EEIS2	2010	detailed
Malawi	IHS	2010	detailed
Rwanda	EICV	2010	detailed
Uganda	UNPS	2010	detailed
Zambia	LCMS	2010	detailed
Niger	LSS	2011	detailed
Sierra leone	SLIHS	2011	detailed
Tanzania	HBS	2011	detailed
South Africa	GHS	2011	detailed
Mauritius	CMPHS	2012	detailed
Nigeria	GHS_2	2012	detailed
Swaziland	HIES	2009	short
The Gambia	IHS	2010	short
Lesotho	HBS	2010	short
Madagascar	EICVM	2010	short
Sao Tome and Principe	IOF	2010	short
Senegal	ESPS	2011	short
Togo	QUIBB	2011	short
Ethiopia	UEUS	2012	short
Ghana	LSS	2012	short

Table 2: Balance

Panel A: Household characteristics, by survey assignment of household

	Households by survey assignment				F-test of equality of coefficients across groups
	Detailed Self- reported	Short Self reported	Detailed Proxy response	Short Proxy response	
Head: female (%)	20.4	26.7	22.3	19.3	0.544
Head: age	48.4	48.7	47.3	48.4	0.882
Head: years of schooling	4.2	4.6	4.5	4.7	0.778
Head: married (%)	74.3	71.6	76.2	81.5	0.277
Household size	6.8	6.2	6.0	6.6	0.046
Share of members less 6 years	16.7	15.4	15.5	16.3	0.876
Share of members 6-15 years	41.2	42.1	41.9	41.0	0.915
Month of interview (1=Jan, 12=Dec)	6.4	6.1	6.3	5.9	0.711
Number of households	113	116	130	135	

Notes: The F-test tests the equality of coefficients across the groups in a regression of each of the household characteristics on group indicators with clustered household standard errors.

Panel B: Individual characteristics, by survey assignment of household

	Detailed Self	Short Self	Detailed Proxy	Short Proxy	F Test for equality of coefficients across groups	
					between all four treatments	between detailed and short
Male	0.46	0.50	0.47	0.47	0.21	0.09
Years of schooling	4.5	4.5	4.3	4.3	0.26	0.79
Age	33.9	34.4	28.9	29.5	0.00	0.43
Married	0.55	0.57	0.48	0.45	0.00	0.86
Number of children below age 6	1.1	1.1	1.1	1.2	0.05	0.86
Number of children age 6-15	1.6	1.5	1.8	1.9	0.00	0.85
Number of old hh members (65+)	0.2	0.2	0.3	0.3	0.68	0.30
Mean community distance to primary school	2.3	2.3	2.3	2.3	0.99	0.87
Mean community distance to secondary school	7.5	7.5	7.7	7.7	0.89	0.96
Mean community distance to all weather road	3.5	3.6	3.7	3.7	0.86	0.94
Number of observations	687	909	723	501		

Table 3: Descriptive statistics

A. By treatment

	Men				Women			
	Detailed	Short	Self	Proxy	Detailed	Short	Self	Proxy
Labor force participation	85%	90%	91%	82%	79%	89%	85%	82%
Wage workers (as % of lfp)	19%	13%	17%	13%	11%	5%	10%	6%
Daily earnings	3987	4781	3634	6255	3912	4388	3552	5367
	(5724)	(4675)	(3492)	(8275)	(6428)	(8422)	(6540)	(8534)
Number of observations	687	723	909	501	785	750	970	565

Standard deviations in parentheses.

B. Wage workers only

	men	women
Daily earnings	4321	4066
Years of schooling	6.60	4.89
Zero years of schooling	15%	37%
1-7 years of schooling	59%	44%
8-11 years of schooling	17%	15%
12-17 years of schooling	8%	4%
Age	33.82	34.09
Married	64%	57%
Number of children below age 6	0.88	1.08
Number of children age 6-15	1.16	1.5
Number of old hh members (65+)	0.10	0.21
Received Detailed questionnaire	57%	68%
Received Short questionnaire	43%	32%
Interviewed self	72%	.72%
Interviewed Proxy	28%	28%

Table 4: Returns to education: treatment and interaction effects in OLS

	Men			Women		
	(1) ln(w)	(2) ln(w)	(3) ln(w)	(4) ln(w)	(5) ln(w)	(6) ln(w)
Years of schooling	0.08***	0.08***	0.04**	0.10***	0.10***	0.08**
	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.04)
Years of schooling X short			**0.06****			**0.05**
			(0.03)			**(0.05)**
Years of schooling X proxy			**0.04**			**0.00**
			(0.03)			**(0.06)**
Short		0.06	-0.35		-0.06	-0.29
		(0.11)	(0.22)		(0.22)	(0.33)
Proxy		0.22	-0.09		0.25	0.23
		(0.14)	(0.28)		(0.19)	(0.37)
D : District dummies	yes	yes	yes	yes	yes	yes
Observations	192	192	192	99	99	99
R-squared	0.44	0.45	0.47	0.37	0.39	0.39

Standard errors, clustered at the household level, in parentheses; *** $p<0.01$, ** $p<0.05$, * $p<0.1$. All regressions include control variables: age, age squared, and a constant

Table 5: Returns to education allowing for nonlinearity and endogeneity

	Men			Women		
	(1)	(2)	(3)	(4)	(5)	(6
	lnw	lnw	lnw	lnw	lnw	lnw
Years of schooling 1 to 7	0.13	0.14	0.12	0.05	0.06	0.0
	(0.101)	(0.102)	(0.107)	(0.080)	(0.083)	(0.0
Years of schooling 8 to 11	0.18*	0.19*	0.15	0.13	0.15*	0.1
	(0.097)	(0.098)	(0.103)	(0.083)	(0.087)	(0.0
Years of schooling 12 to 17	0.19*	0.20**	0.18*	0.18**	0.19**	0.1
	(0.101)	(0.102)	(0.105)	(0.086)	(0.088)	(0.0
Years of schooling 1to 7 X short			0.06			**0.15**
			(0.042)			**(0.0**
Years of schooling 8to 11 X short			0.05			-0.
			(0.036)			(0.0
Years of schooling 12 to 17 X short			**0.06****			**0.0**
			(0.027)			**(0.0**
Years of schooling 1 to7 X proxy			0.00			0.0
			(0.054)			(0.0
Years of schooling 8 to 11 X proxy			0.06			-0.
			(0.044)			(0.0
Years of schooling 12 to 17 X proxy			0.01			0.0
			(0.034)			(0.0
T^j : Treatment variables Short, Proxy	no	yes	yes	no	yes	ye
$\hat{c}$: control function term m/f	yes	yes	yes	yes	yes	ye
D : District dummies	yes	yes	yes	yes	yes	ye
Observations	192	192	192	99	99	9
R-squared	0.486	0.496	0.517	0.453	0.464	0.5

Standard errors, clustered at the household level, in parentheses; *** p<0.01, ** p<0.05, * p<0.1. All regressions
include control variables: age, age squared, mean distance to nearest all season road, and a constant

Table 6: First stage to obtain control function term

	Men	Women
	Years of schooling	Years of schooling
Community mean distance to primary school	-0.21	-0.55**
	(0.203)	(0.244)
Community mean distance to secondary school	-0.17**	-0.15*
	(0.078)	(0.088)
Community mean distance to nearest all-season road	0.05	0.06
	(0.106)	(0.094)
D : District dummies	yes	yes
Observations	192	99
R-squared	0.324	0.391

Standard errors, clustered at the household level, in parentheses; *** p<0.01, ** p<0.05, * p<0.1. All regressions
include control variables: age, age squared, and a constant.

Table 7: Returns to education, allowing for nonlinearity, endogeneity and selection correction

	Men		Women	
	(1)	(2)	(3)	(4)
	Heckman	Heckman - Hotz	Heckman	Heckman - Hotz
	lnw	lnw	lnw	lnw
Years of schooling 1 to 7	0.12	0.12	0.01	0.02
	(0.106)	(0.108)	(0.088)	(0.092)
Years of schooling 8 to 11	0.14	0.15	0.29**	0.21
	(0.103)	(0.105)	(0.126)	(0.128)
Years of schooling 12 to 17	0.14	0.18	0.29*	0.18
	(0.113)	(0.119)	(0.148)	(0.149)
Years of schooling 1 to 7 X short	0.06	0.06	**0.14****	**0.14****
	(0.041)	(0.042)	**(0.065)**	**(0.068)**
Years of schooling 8 to 11 X short	0.05	0.05	-0.03	-0.04
	(0.036)	(0.037)	(0.048)	(0.056)
Years of schooling 12 to 17 X short	**0.05***	**0.06****	**0.08*****	**0.08****
	(0.026)	**(0.027)**	**(0.030)**	**(0.036)**
Years of schooling 1 to 7 X proxy	0.00	0.00	0.03	0.03
	(0.054)	(0.055)	(0.081)	(0.084)
Years of schooling 8 to 11 X proxy	0.06	0.06	-0.04	-0.05
	(0.044)	(0.045)	(0.063)	(0.062)
Years of schooling 12 to 17 X proxy	0.00	0.01	0.10	0.09
	(0.033)	(0.034)	(0.068)	(0.082)
T^j : Treatment variables Short, Proxy	yes	yes	yes	yes
$\hat{c}$: control function term m/f	yes	yes	yes	yes
λ : Mills term m/f	yes	no	yes	no
$h(\hat{p})$: Predicted probability wage worker m/f	no	yes	no	yes
D : District dummies	yes	yes	yes	yes
Observations	192	192	99	99
R-squared	0.521	0.517	0.537	0.517

Standard errors, clustered at the household level, in parentheses; *** p<0.01, ** p<0.05, * p<0.1. All regressions include control variables: age, age squared, mean distance to nearest all season road, and a constant.

Table 8: First stage selection equation

	Men (1) Wage worker Heckman	(2) Wage worker Heckman-Hotz	Women (3) Wage worker Heckman	(4) Wage worker Heckman-Hotz
Years of schooling 1 to 7	0.00	0.00	-0.02	-0.02
	(0.019)	(0.019)	(0.020)	(0.020)
Years of schooling 8 to 11	0.02	0.01	**0.06***	**0.06***
	(0.016)	(0.016)	**(0.021)**	**(0.022)**
Years of schooling 12 to 17	**0.08***	**0.07***	**0.10***	**0.10***
	(0.019)	**(0.019)**	**(0.033)**	**(0.033)**
Short	**-0.23**	**-0.23**	**-0.39***	**-0.39***
	(0.091)	**(0.091)**	**(0.113)**	**(0.113)**
Proxy	-0.16*	-0.16	-0.12	-0.12
	(0.098)	(0.098)	(0.115)	(0.116)
Married	-0.10	-0.10	-0.20	-0.20
	(0.138)	(0.138)	(0.125)	(0.124)
Number of children	-0.06**	-0.06**	-0.04	-0.04
	(0.026)	(0.026)	(0.030)	(0.030)
Community mean distance to primary school		-0.02		-0.00
		(0.026)		(0.026)
Community mean distance to secondary school		0.00		-0.01
		(0.010)		(0.012)
Community mean distance to nearest all-season road	no	yes	no	yes
D : Districts	yes	yes	yes	yes
Observations	1,404	1,404	1,525	1,525

Standard errors, clustered at the household level, in parentheses; *** p<0.01, ** p<0.05, * p<0.1. All regressions include control variables: age, age squared, and a constant

Table 9: Overview of returns to education estimates

	Men →						Women →					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	OLS	Linear estimates after controlling for endogeneity and selection	Non-linear returns	Non-linear returns after controlling for endogeneity	Non-linear returns after controlling for endogeneity and selection Heckman	Non-linear returns after controlling for endogeneity and selection Heckman - Hotz	OLS	Linear estimates after controlling for endogeneity and selection	Non-linear returns	Non-linear returns after controlling for endogeneity	Non-linear returns after controlling for endogeneity and selection Heckman	Non-linear returns after controlling for endogeneity and selection Heckman-Hotz
	Table 4 Col 3	Table A.3 Col 1	Not reported	Table 5 Col 3	Table 7 Col 1	Table 7 Col 2	Table 4 Col 6	Table A.3 Col 2	Not reported	Table 5 Col 6	Table 7 Col 7	Table 7 Col 7
Detailed self	0.04**	0.16*	-0.002 0.03 0.05***	0.12 0.15 0.18*	0.12 0.14 0.14	0.12 0.15 0.18	0.08**	0.05	-0.01 0.15*** 0.16***	0.01 0.17* 0.13	0.01 0.29** 0.29*	0.02 0.21 0.18
Short	0.10***	0.22**	0.05 0.07** 0.11***	0.18 0.20* 0.24**	0.18 0.19* 0.19	0.18 0.20* 0.24*	0.13***	0.04	0.13** 0.12*** 0.17***	0.16 0.15 0.19*	0.15 0.26** 0.37**	0.16 0.17** 0.26**
Difference short - detailed self	0.06**	0.06**	0.05 0.04 0.06**	0.06 0.05 0.06**	0.06 0.05 0.05*	0.06 0.05 0.06**	0.05	-0.01	0.14** -0.03 0.01	0.15** -0.02 0.06**	0.14** -0.03 0.08***	0.14** -0.04 0.08**

Standard errors, clustered at the household level, in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

Annex

Table A.1. Key employment questions in short and detailed questionnaires

Short questionnaire	Detailed questionnaire
	1. During the past 7 days, has [NAME] worked for someone who is not a member of your household, for example, an enterprise, company, the government or any other individual? YES...1 (»3) NO.....2 (question repeated for the past 12 months – question 2)
	3. During the past 7 days, has [NAME] worked on a farm owned, borrowed or rented by a member of your household, whether in cultivating crops or in other farm maintenance tasks, or have you cared for livestock belonging to a member of your household? YES...1 (»5) NO.....2 (question repeated for the past 12 months – question 4)
	5. During the past 7 days, has [NAME] worked on his/her own account or in a business enterprise belonging to he/she or someone in your household, for example, as a trader, shop-keeper, barber, dressmaker, carpenter or taxi driver? YES...1 (»7) NO.....2 (question repeated for the past 12 months – question 6)
1. Did [NAME] do any type of work in the last seven days? Even if for 1 hour. YES...1 (»3) NO.....2 (question repeated for the past 12 months – question 2)	7. CHECK THE ANSWERS TO QUESTIONS 1, 3 AND 5. (WORKED IN LAST 7 DAYS) ANY YES..1 ALL NO.....2 (»37)
3. What is [NAME]'s primary occupation in [NAME]'s main job? (MAIN OCCUPATION IN THE LAST 7 DAYS) a. OCCUPATION b. OCCUPATION CODE	8. What is [NAME]'s primary occupation in [NAME]'s main job? (MAIN OCCUPATION IN THE LAST 7 DAYS) a. OCCUPATION b. OCCUPATION CODE
4. In what sector is this main activity? AGRICULTURE. . . . . 1 MINING/QUARRYING . . . . 2 MANUFACTURING/ PROCESSING. . . . . . . 3 GAS/WATER/ELECTRICITY. . . . . 4 CONSTRUCTION . . . . 5 TRANSPORT. . . . . 6 BUYING AND SELLING . . . . 7 PERSONAL SERVICES. . . . . 8 EDUCATION/HEALTH . . . . 9 PUBLIC ADMINISTRATION. . . . . . . .10 DOMESTIC DUTIES. . . . . .11 OTHER, SPECIFY 12	9. In what sector is this main activity? AGRICULTURE. . . . . 1 MINING/QUARRYING . . . . 2 MANUFACTURING/ PROCESSING. . . . . . . 3 GAS/WATER/ELECTRICITY. . . . . 4 CONSTRUCTION . . . . 5 TRANSPORT. . . . . 6 BUYING AND SELLING . . . . 7 PERSONAL SERVICES. . . . . 8 EDUCATION/HEALTH . . . . 9 PUBLIC ADMINISTRATION. . . . . . . .10 DOMESTIC DUTIES. . . . . .11 OTHER, SPECIFY 12

Table A.2: Planned and actual survey assignments

	Household survey assignment				
	Detailed self-reported	Detailed proxy response	Short self-reported	Short proxy	Total
Households					
Number (planned = actual)	336	336	336	336	1344
Percent with one adult 15+	14.0	12.2	14.6	11.9	
Percent with one member 10+	9.8	9.2	10.7	10.7	
Planned individual assignment, if every household has at least 3 members over 10 years of age, and at least one member over 15 years.^					
Detailed self-reported	672	336	0	0	1008
Detailed proxy response	0	672	0	0	672
Short self-reported	0	0	672	336	1008
Short proxy planned	0	0	0	672	672
Planned individual assignment, given assumption about household composition^ #					
Detailed self-reported	672	336	0	0	1008
Detailed proxy response	0	504	0	0	504
Short self-reported	0	0	672	336	1008
Short proxy planned	0	0	0	504	504
Actual individual assignment					
Detailed self-reported	606	336	0	0	942
Detailed proxy response	32	498	0	0	530
Short self-reported	0	0	601	336	937
Short proxy	0	0	35	501	536
Total actual number of individuals					2,945
Numbers of observations for different groups					
Detailed	638	834	0	0	**1472**
Short	0	0	636	837	**1473**
Self	606	336	601	336	**1879**
Proxy	32	498	35	501	**1066**

^ Assuming that each household has at least 2 persons age 10+ to be randomly selected for self-report.

Assuming that each household has one member 15+ and an average of 2.5 household members 10+years per household. Thus, there are 1.5 *336 other members to be reported on by proxy.

Table A.3: Returns to education, after controlling for endogeneity (but not nonlinearity)

	Men (1) lnw	Women (2) lnw
Years of schooling	0.16*	0.05
	(0.096)	(0.088)
Years of schooling X short	0.06**	0.04
	(0.027)	(0.047)
Years of schooling X proxy	0.03	-0.00
	(0.034)	(0.056)
T^j : Treatment variables Short, Proxy	yes	yes
$\hat{c}$: control function term m/f	yes	yes
$h(\hat{p})$: Predicted probability wage worker m/f	yes	yes
D : District dummies	yes	yes
Observations	192	99
R-squared	0.492	0.417

Standard errors, clustered at the household level, in parentheses; *** $p<0.01$, ** $p<0.05$, * $p<0.1$. All regressions include control variables age, age squared, mean distance to nearest all season road, and a constant